Abstract

Peer-to-peer (P2P) databases are becoming prevalent on the Internet for distribution and sharing of documents, applications, and other digital media. The problem of answering large-scale ad hoc analysis queries, for example, aggregation queries, on these databases poses unique challenges. Exact solutions can be time consuming and difficult to implement, given the distributed and dynamic nature of P2P databases. In this paper, we presented novel
sampling-based techniques for approximate answering of ad hoc aggregation queries in such databases. Computing a high-quality random sample of the database efficiently in the P2P environment is complicated due to several factors: the data is distributed (usually in uneven quantities) across many peers, within each peer, the data is often highly correlated, and, moreover, even collecting a random sample of the peers is difficult to accomplish. To counter these problems, the proposed approach will use an approach based on random walks of the P2P graph, as well as block-level sampling techniques.

References


Index Terms
Computer Science
Engineering and Technology

Keywords
Peer-to-peer Network  Query Processing  Distributed Databases